PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C07B 61/00, C07K 1/04, C12Q 1/68, G01N 33/543, 33/551, 33/552, 33/544

A1

(11) International Publication Number: WO 00/09464

(43) International Publication Date: 24 February 2000 (24.02.00)

(21) International Application Number: PCT/US99/18600

(22) International Filing Date: 16 August 1999 (16.08.99)

(30) Priority Data:
60/096,820 17 August 1998 (17.08.98) US

(71) Applicant: PHYLOS, INC. [US/US]; 128 Spring Street, Lexington, MA 02421 (US).

(72) Inventor: LOHSE, Peter; 28 Skahan Road, Belmont, MA 02178 (US).

(74) Agent: ELBING, Karen; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110–2214 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: IDENTIFICATION OF COMPOUND–PROTEIN INTERACTIONS USING LIBRARIES OF PROTEIN–NUCLEIC ACID FUSION MOLECULES

(57) Abstract

Disclosed herein is a method for detecting a compound–protein interaction, involving: (a) providing a compound library in which each member of the compound library is immobilized on a solid support; (b) contacting each member of the immobilized compound library in a single reaction chamber with each member of a protein–nucleic acid fusion library under conditions which allow the formation of compound–fusion complexes; (c) isolating the immobilized compound–fusion complexes; and (d) detecting a compound–fusion complex as an indication that the protein of the fusion interacts with the compound. In preferred embodiments, the protein is identified by reading the nucleic acid portion of the fusion, and the compound is identified by reading a detectable tag bound to either the compound or the solid support.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

IDENTIFICATION OF COMPOUND-PROTEIN INTERACTIONS USING LIBRARIES OF PROTEIN-NUCLEIC ACID FUSION MOLECULES

Background of the Invention

In general, the invention features screening methods involving nucleic acid-protein fusions.

Screening is considered to be an efficient tool to identify binding interactions between proteins and small molecule compounds derived from large pharmaceutically-based collections, new synthetic approaches such as combinatorial chemistry, or natural sources (TIBTECH, vol. 13, p. 115, 1995). However, the multidisciplinary nature of most screening techniques poses significant challenges. The most important challenge of such techniques is maintaining a ready supply of materials for the screen. Screening of small molecule compound libraries with different protein targets requires sufficient amounts of compound. Alternatively, screening of large compound libraries (for example, having 10^6 members or greater) requires large amounts of recombinant protein. Another challenge is to operate the screen rapidly and cost effectively. Screening of compound libraries with different protein targets is generally time consuming if carried out in a sequential fashion.

Lately, a method has been described for the isolation of proteins with desired properties out of a pool of proteins (Szostak et al., Selection of Proteins Using RNA-Protein Fusions, U.S.S.N. 09/007,005, January 14, 1998, and U.S.S.N. 09/247,190, February 9, 1999; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302). This technique is accomplished by means of protein-RNA fusion molecules where each protein is covalently linked to its encoding RNA. The protein-RNA fusion technology may be used to screen cDNA libraries and to clone new genes on the basis of protein-protein interactions (see, for example, Szostak et al., Selection of

Proteins Using RNA-Protein Fusions, U.S.S.N. 09/007,005, January 14, 1998, and U.S.S.N. 09/247,190, February 9, 1999).

Summary of the Invention

The purpose of the present invention is to efficiently identify protein-compound binding interactions (and, particularly, protein-small molecule interactions) by screening small molecule compounds with libraries of protein-nucleic acid fusions (for example, protein-RNA fusions) in a parallel fashion, thus providing a catalogue of small molecule-protein pairs.

Accordingly, in a first aspect, the invention features a method for detecting a compound-protein interaction, the method involving: (a) providing a compound library in which each member of the compound library is immobilized on a solid support; (b) contacting each member of the immobilized compound library in a single reaction chamber with each member of a protein-nucleic acid fusion library under conditions which allow the formation of compound-fusion complexes; (c) isolating the immobilized compound-fusion complexes; and (d) detecting the compound-fusion complex as an indication that the protein of the fusion interacts with the compound.

In preferred embodiments, the protein-nucleic acid fusion is either a protein-RNA fusion, a protein-DNA fusion, or a protein fused to a DNA-RNA hybrid; the solid support is a bead; each bead is coded with a unique detectable label; the compound of the complexed protein-nucleic acid fusion is identified by the unique detectable label associated with the bead; the detectable label is a peptide label, a nucleic acid label, a chemical label, a fluorescent label, or a radio frequency tag; the solid support is a chip and the compound library is immobilized on the chip in an addressable array; each member of the protein-nucleic acid fusion library is detectably labeled; the compound-fusion complex, or the components thereof, are recovered by release from the solid support; the

method further involves recovering the protein-nucleic acid fusion from the solid support and identifying the protein; the identity of the protein is determined from the sequence of the nucleic acid portion of the protein-nucleic acid fusion; and the compound is a small molecule.

In a related aspect, the invention features a method for detecting a compound-protein interaction, the method involving: (a) providing a compound immobilized on a solid support; (b) contacting the immobilized compound with a protein-nucleic acid fusion library under conditions which allow the fusion to bind to the compound; and (c) detecting a bound protein-nucleic acid fusion as an indication that the protein of the protein-nucleic acid fusion interacts with the compound.

In preferred embodiments, the protein-nucleic acid fusion is either a protein-RNA fusion, a protein-DNA fusion, or a protein fused to a DNA-RNA hybrid; the protein-nucleic acid fusion is detectably labeled and the interaction is indicated by the association of the detectable label with the solid support; the bound protein-nucleic acid fusion is recovered by release from the solid support; the method further involves recovering the protein-nucleic acid fusion from the solid support and identifying the protein; the identity of the protein is determined from the sequence of the nucleic acid portion of the protein-nucleic acid fusion; the solid support is a column, glass slide, chip, or bead; and the compound is a small molecule.

As used herein, by a "library" is meant a collection of at least two molecules (for example, molecules such as compounds or protein-nucleic acid fusions). A compound library preferably includes at least 10^2 or 10^3 members, and, more preferably, at least 10^4, 10^5, or 10^6 members. A protein-nucleic acid library (for example, a protein-RNA library) preferably includes at least 10^2 or 10^3 members, more preferably, at least 10^4, 10^5, or 10^6 members, and, most preferably, at least 10^{10} or 10^{12} members.

By a "DNA-RNA hybrid" is meant a DNA strand hybridized to a complementary RNA strand. Typically, the DNA strand is generated by reverse transcription of the RNA molecule.

By "addressable array" is meant a fixed pattern of immobilized objects on a solid surface in which the identity of the objects is known or can be readily determined.

By a "small molecule" is meant a compound with a molecular weight of less than or equal to 10,000 Daltons, preferably, less than or equal to 1000 Daltons, and, most preferably, less than or equal to 500 Daltons.

The present invention provides a number of advantages. For example, the present methods reduce the amount of material required for a screen. In standard screens, considerable amounts of protein and small molecule compounds are required because each compound is screened with a single protein in a spatially segregated chamber. A library of protein-nucleic acid fusion molecules, however, can be screened for binding interactions with small molecule compounds in the same reaction chamber in a parallel fashion. In addition, the protein target need not be cloned, overexpressed, or isolated, but rather is screened as a protein-nucleic acid fusion molecule and identified by its coding nucleic acid. Moreover, material costs may be further reduced by miniaturization, which is facilitated by the present methods and is limited solely by the choice of detection method for the identification of small molecule-fusion complexes.

In addition, the present invention provides advantages in terms of the time required to carry out a compound screen. In particular, the methods described herein accelerate the identification of ligands (for example, small molecule ligands) by screening a library of protein targets with a library of potential ligands in a parallel fashion. In contrast to standard screens, where a

small compound library is screened for binding to different proteins in a sequential fashion, small molecule compounds may be screened, in the present techniques, with a library of protein-nucleic acid fusions in a single assay. Consequently, the present invention facilitates the screening of members of a library of small molecule compounds for binding to the members of a library of proteins in a highly efficient manner.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Drawings

FIGURE 1 is a schematic illustration of an exemplary approach to screening a compound immobilized on a solid support with a library of protein-nucleic acid fusions.

FIGURE 2 is a schematic illustration of an exemplary approach to screening a library of compounds immobilized to beads with a library of protein-nucleic acid fusions.

FIGURE 3 is a schematic illustration of an exemplary approach to screening an addressable array of compounds immobilized on a microchip with a library of protein-nucleic acid fusions.

FIGURE 4 is a graph illustrating compound binding to an RNA-protein fusion on a bead solid support.

Detailed Description

The methods of the present invention facilitate the efficient identification of protein-compound (and, preferably, protein-small molecule) binding interactions by screening such compounds with libraries of protein-nucleic acid fusions (for example, protein-RNA fusions), thus providing a catalog of compound-protein pairs. If desired, libraries of compounds may be

screened against libraries of protein-nucleic acid fusions in a single screen. In preferred embodiments, either the compounds or the fusions are immobilized on a solid support (for example, a bead, chip, glass slide, or column) to simplify the screen and/or result readings. In addition, to facilitate the identification of compound-protein pairs, the compound (or the solid support to which it is immobilized) may be tagged with a detectable label characteristic of that particular compound or compound family.

Any compound may be screened by the methods of the invention, although small molecules represent preferred targets for screening.

These and other aspects of the invention are now described in more detail below. These examples are provided for the purpose of illustrating the invention, and should not be construed as limiting.

Screening Assays

As discussed above, screening of compounds against protein-nucleic acid fusions (for example, protein-RNA fusions) may be carried out in a number of different formats. One particular format is illustrated in Figure 1. By this approach, a single compound is immobilized on a column or any other solid surface using any one of a variety of standard methods. The solid phase-bound small molecule compound is then incubated with screening buffer containing BSA or another inert protein to reduce non-specific binding. Next, the buffer solution is removed, and the solid phase presenting the compound is incubated with a solution of a protein-nucleic acid fusion library, followed by washes with screening buffer to remove non-specifically bound fusion molecules. Specifically bound protein-nucleic acid fusions are then eluted (for example, by affinity elution using buffer containing free small molecule compound). "Reading" the nucleic acid (for example, RNA) portion of the eluted fusion molecules provides an identification of the protein that bound the

small molecule compound. Such a "reading" may be carried out as described below.

Alternatively, multiple compounds may be screened simultaneously against multiple protein-nucleic acid fusions. Two exemplary formats for carrying out this type of screen are shown in Figures 2 and 3. In these formats, an encoded (addressable) library of small molecules is immobilized on beads or any other surface, such as a chip. The solid phase-bound library is then incubated with screening buffer containing BSA or another inert protein to reduce non-specific binding. Subsequently, the buffer solution is drained, and the small molecule compound library is incubated with a fusion library, followed by washes with screening buffer to remove non-specifically bound molecules. Protein-nucleic acid fusion molecules specifically binding to small molecules are then detected or, if a bead format is utilized, sorted and collected. A reading code (or tag or address) is used to identify the small molecule compound, and reading of the nucleic acid portion of bound fusion molecules is used to identify the protein (as described below).

Protein-nucleic acid fusion molecules of different genotypes and different phenotypes can sometimes bind to the same small molecule compound. If desired, therefore, the bound fraction of fusion molecules may be collected, amplified, and reincubated with an identified ligand under more stringent conditions (e.g., a lower concentration of protein-nucleic acid fusion). This process may be repeated any number of times, allowing for the isolation of a receptor with any desired ligand affinity (for example, selection for a receptor having the highest affinity).

In addition, once identified, a binding interaction between a solid phase-bound compound and a fusion molecule may be confirmed or analyzed by addition of free ligand or free protein to a compound-fusion complex in a standard binding assay.

The present screens may be used to identify unknown compound-protein interactions or may be exploited in circumstances where some general knowledge of an interaction (for example, between a ligand and a receptor) is available. In the latter case, biased libraries may be used for screening. Such libraries may contain particular classes of compounds (or proteins) or modifications of a single compound (or protein). In general, the biasing element tends to increase the average affinity of a ligand for a target receptor and to orient the ligand in a uniform way (see, for example, Chen et al., JACS (1993) vol. 115, p. 12591-12592). This type of approach facilitates the identification, for example, of ligands which bind to a receptor at a targeted site.

Preparation of Protein-Nucleic Acid Fusions

As discussed above, the present techniques may be applied to any population of protein-nucleic acid fusions, including protein-RNA fusions, protein-DNA fusions, and fusions between proteins and hybrid DNA-RNA molecules.

For use in the methods described herein, random libraries of protein-RNA fusion molecules may be prepared, for example, as described in Szostak et al., Selection of Proteins Using RNA-Protein Fusions, U.S.S.N. 09/007,005, January 14, 1998, and U.S.S.N. 09/247,190, February 9, 1999; Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302; or Kuimelis et al., Addressable Protein Arrays, U.S.S.N. 60/080,686, April 3, 1998, and U.S.S.N. 09/282,734, March 31, 1999). Alternatively, libraries of cellular RNA-protein fusion molecules may be prepared from mRNAs or cDNAs that lack 3'-untranslated regions, for example, as described in Lipovsek et al. (Methods for Optimizing Cellular RNA-Protein Fusion Formation, U.S.S.N. 60/096,818, August 17, 1998) and Hammond et al. (Methods for Producing Nucleic Acids Lacking 3'-Untranslated Regions and Optimizing

Cellular RNA-Protein Fusion Formation, U.S.S.N. _____, August 17, 1999).

To label such protein-RNA fusions, any standard labeling method and any detectable label (including, for example, radioactive, fluorescent, and chemiluminescent labels) may be utilized. If desired, fusions may be radioactively labeled by generating the fusion or fusion components in the presence of radioactive amino acids (for example, ^{35}S- or ^{14}C-labeled amino acids) or radioactive nucleotides (for example, ^{35}S- or ^{32}P-labeled nucleotides). Alternatively, fusion molecules may be fluorescently labeled. In one particular example, the DNA linker (for example, the $dA_{27}dCdCP$ linker described in Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302) may be modified with a fluorescein phosphoramidite marker (Glen Research, Sterling, VA), and this linker used for the synthesis of fluorescent protein-RNA fusions. In yet another alternative, protein-RNA fusions prepared according to the method of Roberts & Szostak (Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302; and Selection of Proteins Using RNA-Protein Fusions, U.S.S.N. 09/007,005, January 14, 1998, and U.S.S.N. 09/247,190, February 9, 1999) or cellular RNA-protein fusions prepared according to the method of Lipovsek et al. (Methods for Optimizing Cellular RNA-Protein Fusion Formation, U.S.S.N. 60/096,818, August 17, 1998) or Hammond et al. (Methods for Producing Nucleic Acids Lacking 3'-Untranslated Regions and Optimizing Cellular RNA-Protein Fusion Formation, U.S.S.N. _____, August 17, 1999) may be labeled by base pairing the fusion to a fluorescently-labeled oligonucleotide (for example, base pairing a fluorescent poly-dT oligonucleotide to the $dA_{27}dCdCP$ linker).

Alternatively, protein-DNA fusions may also be labeled using similar techniques. Such protein-DNA fusions may be generated as described, for example, in Lohse et al., DNA-Protein Fusions and Uses Thereof, U.S.S.N.

60/110,549, December 2, 1998. In yet another alternative, the above labeling techniques may be used for fusions of proteins to hybrid DNA-RNA portions (i.e., one strand of each). Such hybrid fusions are generated, for example, by subjecting a RNA-protein fusion to a step of reverse transcription using standard techniques.

Preparation of Compounds

For carrying out the screening methods of the invention, any compound library may be utilized. Such libraries may be derived from natural products, synthetic (or semi-synthetic) extracts, or chemical libraries according to methods known in the art. Those skilled in the field of drug discovery and development will understand that the precise source of compounds is not critical to the screening procedure(s) of the invention. Examples of natural compound sources include, but are not limited to, plant, fungal, prokaryotic, or animal sources, as well as modification of existing compounds. Numerous methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Synthetic compound libraries may be obtained commercially or may be produced according to methods known in the art. Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

In certain methods of the invention, interacting compounds are identified as a result of a detectable label, or "tag," bound to either the compound or its associated solid support (for example, bead). A coded library of small molecule compounds may be prepared on beads as described, for example, in Combs et al., JACS (1996) vol. 118, p. 287-288. In addition, a number of encoding schemes are available, including peptide and nucleic acid codes (Kerr et al., JACS (1993) vol. 115, p. 2529-2531; and Brenner & Lerner,

Proc. Natl. Acad. Sci. USA (1992) vol. 89, p. 5381-5383); chemical tags (Ohlmeyer et al., Proc. Natl. Acad. Sci. USA (1993) vol. 90, p. 109222-10926; and Maclean et al., Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 2805-2810); fluorophore tags (Yamashita & Weinstock (SmithKline Beecham Corporation), WO95/32425 (1995); and Sebestyen et al., Pept. Proc. Eur. Pept. Symp. 22nd 1992 (1993), p. 63-64); and radio frequency tags (Nicolaou et al., Angew. Chem. Int. Ed. Engl. (1995) vol. 34, p. 2289-2291; and Moran et al., JACS (1995) vol. 117, p. 10787-10788). Such labels may be read as described in the references above.

Alternatively, an addressable library of compounds (for example, small molecule compounds) may be prepared on a solid surface, such as a chip surface. A variety of techniques are available for immobilizing compounds on a chip surface, and any may be utilized. Preferable techniques include photolithography (Affymetrix, Santa Clara, CA), mechanical microspotting (Schena et al., Science (1995) vol. 270, p. 467-470; Synteni, Fremont, CA) and ink jetting (Incyte Pharmaceuticals, Palo Alto, CA; and Protogene, Palo Alto, CA).

Identification of Compound-Fusion Interactions

To identify interactions between compounds (for example, coded compounds) and protein-nucleic acid fusions, any method may be utilized which provides a means for detecting a label associated with the compound or fusion or, if appropriate, for isolating and determining the identity or "address" of the compound-fusion pair.

In one particular example, compound-protein pairs (for example, small molecule-protein pairs) may be isolated and identified on beads. To detect a label associated with a bead, the bead resin is preferably plated out, followed by scanning, for example, for a fluorescent or radioactive label (using,

for example, a Phosphorimager to detect a radioactive label). Protein-nucleic acid fusion molecules binding to small molecules presented on a bead may be isolated by physically sorting the beads. Alternatively, beads bound to fluorescently labeled fusion molecules may be sorted on a fluorescence activated cell sorter (FACS). Selected beads may be individually and spatially separated (for example, into the wells of a 96-well microtiter plate). For RNA-protein fusions, molecules bound to individual beads may then be identified by reverse transcription of the RNA portion, followed by sequencing of the DNA as described by Roberts & Szostak (Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302) and Szostak et al. (Selection of Proteins Using RNA-Protein Fusions, U.S.S.N. 09/007,005, January 14, 1998, and U.S.S.N. 09/247,190, February 9, 1999). The tag coding for the compound (for example, the small molecule compound) on each individual bead may be read as described above.

Alternatively, ligand-receptor pairs on a chip surface may be detected by scanning the chip surface for radioactivity or fluorescence. The address of the interacting pair on the chip reveals the identity of the compound (for example, the small molecule compound). The fusion molecule may be picked from the chip surface using an addressable microcollector or any other standard method (see, for example, Kuimelis et al., Addressable Protein Arrays, U.S.S.N. 60/080,686, April 3, 1998, and U.S.S.N. 09/282,734, March 31, 1999). The retrieved fusion molecule may then be identified by characterizing the nucleic acid portion of the fusion as described above.

Compound Screening Utilizing a Bead Format

As described above, compounds may be immobilized on a bead solid support and used to screen for protein-nucleic acid fusions, and specifically for RNA-protein fusions, which are capable of interacting with the compound. In one particular working example of this approach, the dihydrofolate reductase

(DHFR) gene was cloned out of a human liver cDNA library (Maxim Biotech, South San Francisco, CA). The construct contained the entire DHFR gene with an added C-terminal DYKDDDDK-ASA peptide tag (SEQ ID NO: 1). RNA-protein fusions of DHFR were prepared by PCR amplification of the DHFR coding sequence followed by fusion formation as described in Roberts & Szostak (Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302) and Szostak et al. (Selection of Proteins Using RNA-Protein Fusions, U.S.S.N. 09/007,005, January 14, 1998, and U.S.S.N. 09/247,190, February 9, 1999). The fusions were purified using oligo-dT-cellulose affinity chromatography (Edmonds et al., Proc. Natl. Acad. Sci. USA (1971) vol. 68:1336) and reverse transcribed with Superscript II reverse transcriptase according to the manufacturer's instructions. 100 fmol of DHFR fusion in 10 µL 1 X buffer (Phosphate buffered saline, 1 M NaCl, 1 mg/ml BSA, 0.1 mg/ml sheared DNA, 1% v/v Triton X-100) was combined with 10 µL pre-equilibrated methotrexate-agarose (as described in Kaufman, Methods Enzymol. (1974) vol. 34:272-81) in a 500 µL eppendorf tube. The slurry was incubated for 30 minutes at ambient temperature with mixing every 5 minutes. The slurry was then centrifuged for 1 minute at 3000 rpm in an eppendorf microfuge. The liquid was removed, and the methotrexate-agarose was washed 3 times with 500 µL of 1 x buffer. The fusions were then eluted by incubation of the methotrexate-agarose in 50 µL 30 µM methotrexate for 30 minutes at 37°C.

The results of this interaction assay are shown in Figure 4. In this figure, the percent of total fusion molecules was monitored by measuring ^{35}S-methionine label incorporated into the fusions during the translation step. As indicated, the third wash contained no significant amount of fusion molecules. In addition, of the total amount of fusion included within the matrix, 86% flowed through the bead column, and the other 14% was efficiently eluted with methotrexate.

Use

The present methods provide an efficient means for screening either small or large libraries for compound-protein binding interactions. In addition, these methods may be utilized to screen protein-nucleic acid fusions against one compound or against a library of compounds.

Commercial uses for screening a library of fusion molecules against a single compound include, without limitation, identification of a protein binder for a desired small molecule from a random pool of fusion molecules, rationalization of the mechanism of action of a given drug by isolating the cellular target from a pool of cellular mRNA-protein fusion molecules (or a pool of the DNA-protein fusion or hybrid fusion derivatives), and rationalization of the side effect profile of a given drug by isolating most or all target proteins from a pool of cellular mRNA-protein (or DNA-protein or hybrid-protein) fusion molecules, leading to an improved drug with reduced side effects.

Uses for screening a library of fusion molecules against an encoded (addressable) library of compounds include, without limitation, screening a library of small molecule compounds with a library of nucleic acid-protein (for example, cellular mRNA-protein) fusion molecules for potential new lead compounds (for example, ligands or enzyme inhibitors), screening a library of nucleic acid-protein (for example, cellular mRNA-protein) fusion molecules with a library of small molecule compounds for potential targets (for example, receptors or enzymes), and mapping of binding interactions between the members of a protein library and the members of a small molecule compound library, thus providing a catalogue of ligand-protein pairs.

All patents and publications mentioned herein are hereby incorporated by reference.

Other embodiments are within the claims.

What is claimed is:

Claims

1. A method for detecting a compound-protein interaction, said method comprising:

(a) providing a compound library in which each member of said compound library is immobilized on a solid support;

(b) contacting each member of said immobilized compound library in a single reaction chamber with each member of a protein-nucleic acid fusion library under conditions which allow the formation of compound-fusion complexes;

(c) isolating said immobilized compound-fusion complexes; and

(d) detecting said compound-fusion complex as an indication that the protein of said fusion interacts with said compound.

2. The method of claim 1, wherein said protein-nucleic acid fusion is a protein-RNA fusion, a protein-DNA fusion, or a protein fused to a DNA-RNA hybrid molecule.

3. The method of claim 1, wherein said solid support is a bead.

4. The method of claim 3, wherein each said bead is coded with a unique detectable label.

5. The method of claim 4, wherein the compound of said complexed protein-nucleic acid fusion is identified by said unique detectable label associated with said bead.

6. The method of claim 4, wherein said detectable label is a peptide

label, a nucleic acid label, a chemical label, a fluorescent label, or a radio frequency tag.

7. The method of claim 1, wherein said solid support is a chip and said compound library is immobilized on said chip in an addressable array.

8. The method of claim 7, wherein each member of said protein-nucleic acid fusion library is detectably labeled.

9. The method of claim 1, wherein said compound-fusion complex, or the components thereof, are recovered by release from said solid support.

10. The method of claim 1, wherein said method further comprises recovering said protein-nucleic acid fusion from said solid support and identifying said protein.

11. The method of claim 10, wherein the identity of said protein is determined from the sequence of the nucleic acid portion of said protein-nucleic acid fusion.

12. The method of claim 1, wherein said compound is a small molecule.

13. A method for detecting a compound-protein interaction, said method comprising:

(a) providing a compound immobilized on a solid support;

(b) contacting said immobilized compound with a protein-nucleic acid fusion library under conditions which allow said fusion to bind to said

compound; and

(c) detecting a bound protein-nucleic acid fusion as an indication that the protein of said protein-nucleic acid fusion interacts with said compound.

14. The method of claim 13, wherein said protein-nucleic acid fusion is a protein-RNA fusion, a protein-DNA fusion, or a protein fused to a DNA-RNA hybrid molecule.

15. The method of claim 13, wherein said protein-nucleic acid fusion is detectably labeled and said interaction is indicated by the association of said detectable label with said solid support.

16. The method of claim 13, wherein said bound protein-nucleic acid fusion is recovered by release from said solid support.

17. The method of claim 13, wherein said method further comprises recovering said protein-nucleic acid fusion from said solid support and identifying said protein.

18. The method of claim 17, wherein the identity of said protein is determined from the sequence of the nucleic acid portion of said protein-nucleic acid fusion.

19. The method of claim 13, wherein said solid support is a column, glass slide, chip, or bead.

20. The method of claim 13, wherein said compound is a small molecule.

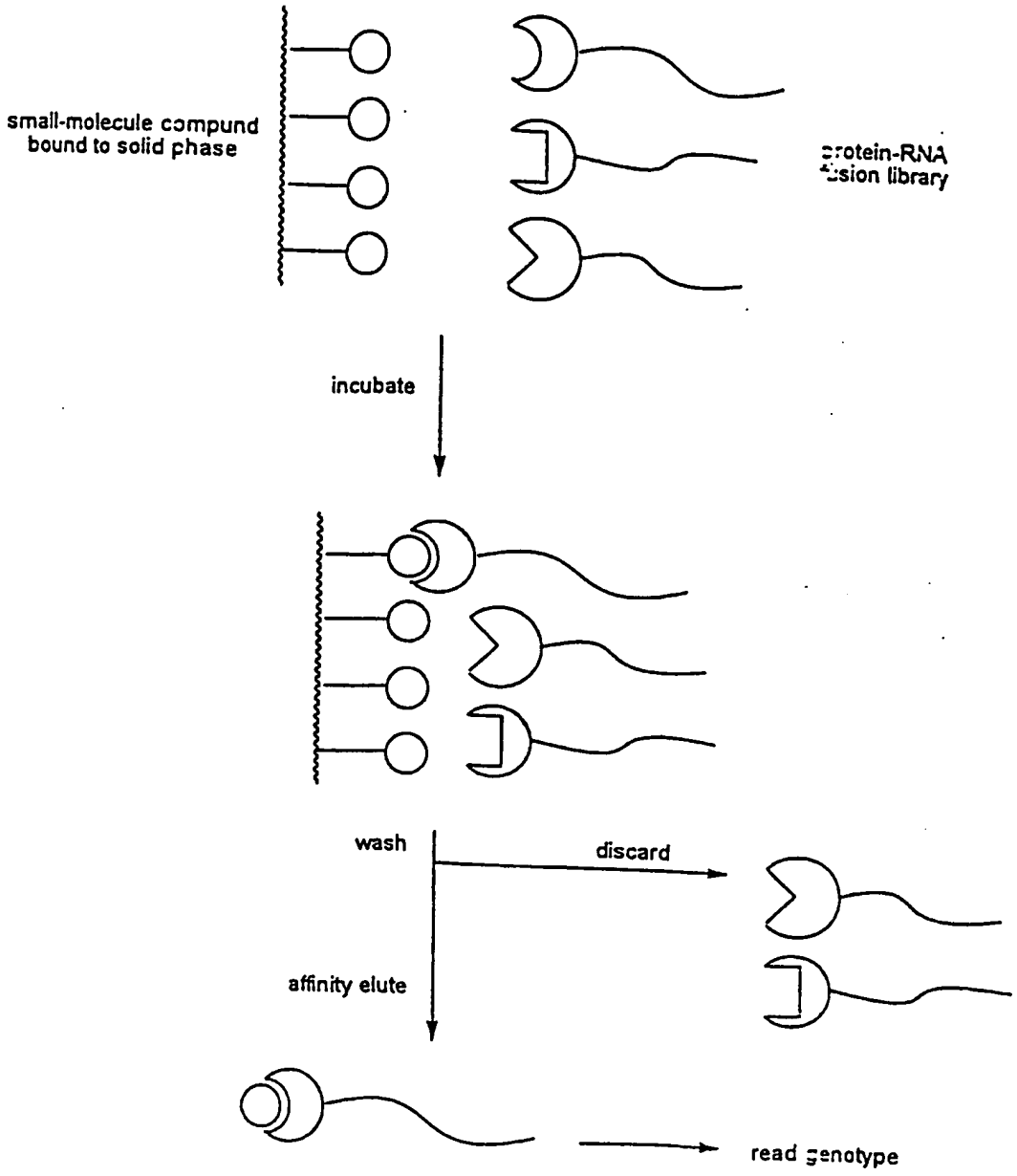
small-molecule compund
bound to solid phase
protein-RNA
fusion library
incubate
wash
discard
affinity elute
read genotype

FIGURE 1

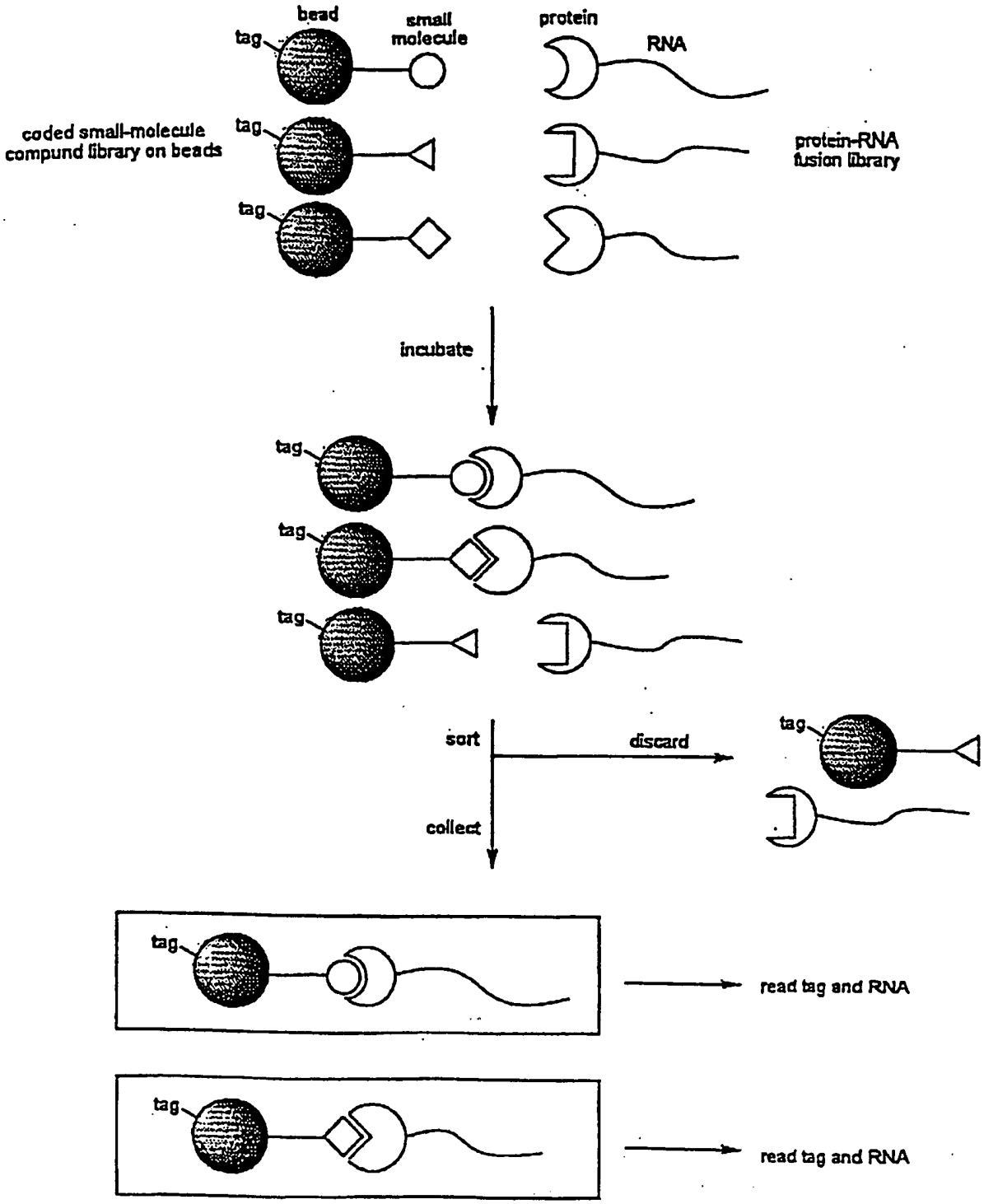
bead
tag
small molecule
protein
RNA
coded small-molecule compund library on beads
protein-RNA fusion library
incubate
sort
discard
collect
read tag and RNA
read tag and RNA

FIGURE 2

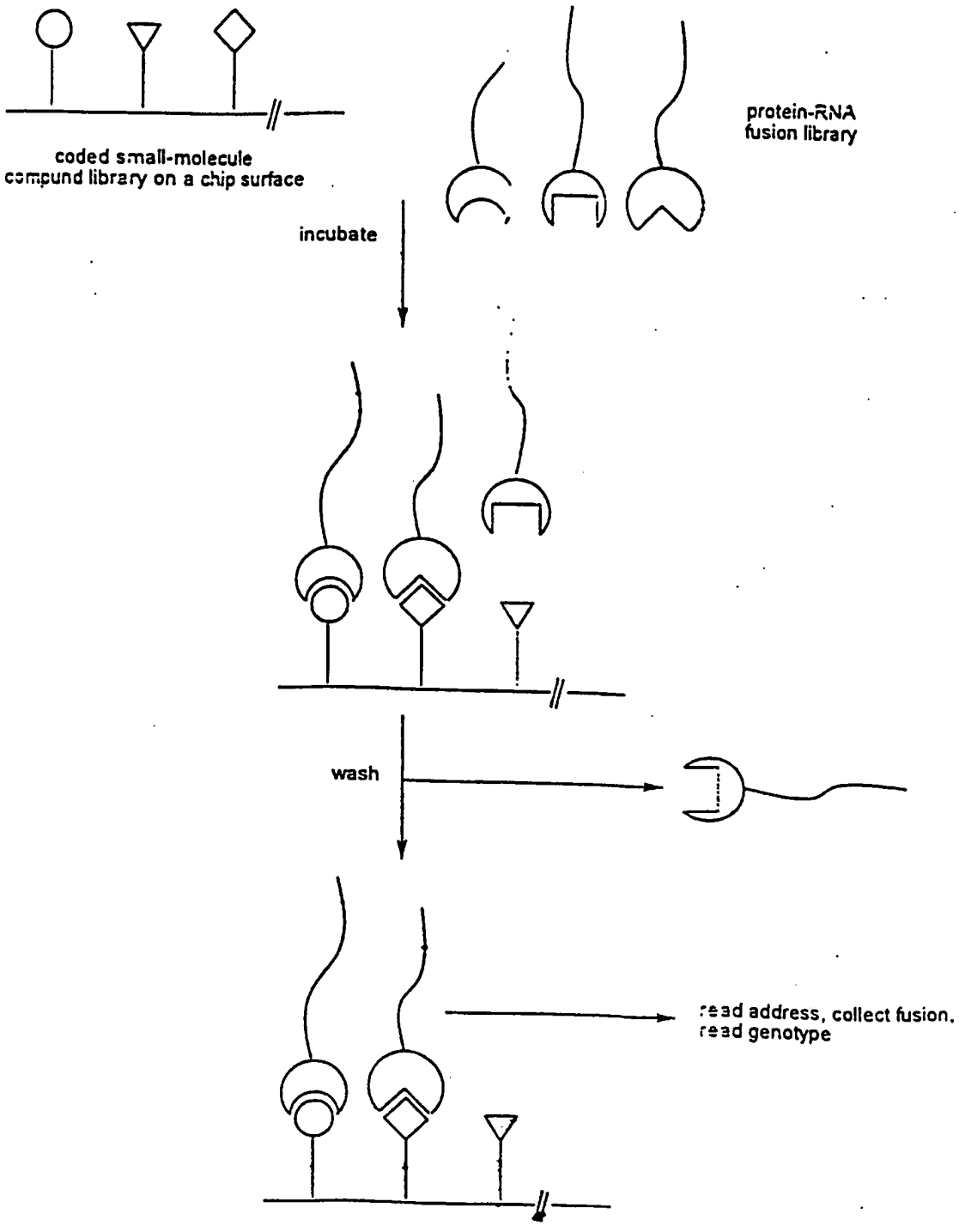
coded small-molecule
compound library on a chip surface
protein-RNA
fusion library
incubate
wash
read address, collect fusion,
read genotype

FIGURE 3

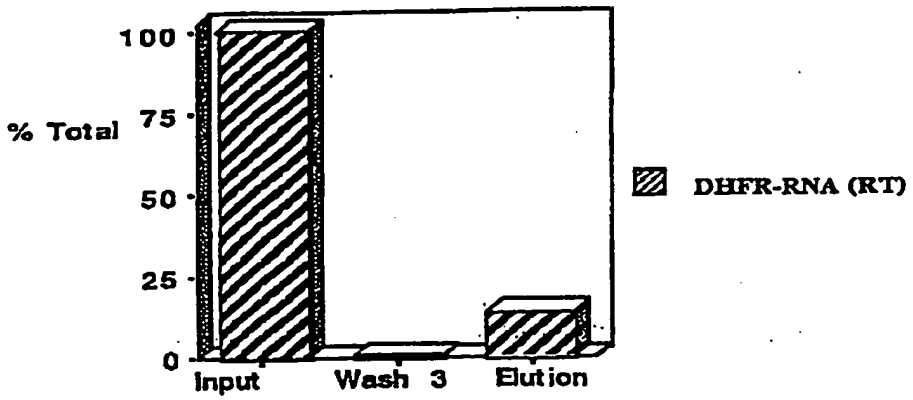

FIGURE 4

SEQUENCE LISTING

<110> Phylos, Inc.

<120> IDENTIFICATION OF COMPOUND-
PROTEIN INTERACTIONS USING LIBRARIES
OF PROTEIN-NUCLEIC ACID FUSION MOLECULES

<130> 50036/017WO2

<150> 60/096,820
<151> 1998-08-17

<160> 1

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic affinity tag

<400> 1

Asp	Tyr	Lys	Asp	Asp	Asp	Asp	Lys	Ala	Ser	Ala
1				5				10		

INTERNATI NAL SEARCH REPORT

International application No.
PCT/US99/18600

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :Please See Extra Sheet.
US CL : 435/6; 436/518, 523, 524, 527, 528
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6; 436/518, 523, 524, 527, 528
435/ Digest 2, Digest 9 , Digest 14, Digest 15 digest 17

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,751,629 A (NOVA et al) 12 May 1998, see the entire document, especially the abstract and column 13 lines 23-28 which discuss the use of chimeric peptide-oligonucleotides.	1-10, 12-17, 19-20
Y	WO 92/00091 A1 (BIOLIGAND, INC.) 09 January 1992, see the entire document, especially the abstract and note that it teaches the use of bio-oligomer which may be a combination of peptides and nucleotides.	1-20
Y	US 5,770,455 A (CARGILL et al) 23 June 1998, see the entire document, note the use of RF tags.	6, 7, 19
Y,P	US 5,565,324 A (STILL et al) 15 October 1998, see the entire document, note the encoding with tags.	1-10, 12-17, 19-20

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
17 DECEMBER 1999	12 JAN 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer JOSEPH W. RICIGLIANO Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/18600

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,424,186 A (FODOR et al) 13 June 1995, se entire document, note the addressable arrays.	7
Y	NEEDLES et al. Generation and screening of an oligonucleotide-encoded synthetic peptide library. Proc. Natl. Acad. Sci. USA. November 1993, Vol. 90, pages 10700-10704, see the entire document.	11, 18

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/18600

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (7):

C07B 61/00L; C07K 1/04C; C12Q 1/68; G01N 33/543, 33/551, 33/552, 33/544

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

WEST: Derwent and US pat files
bio-oligomers, chimera, nucleotide, oligonucleotide, peptide, amino acid library combinatorial screening, peptide-oligonucleotide